Subject CT4 — Models
Core Technical

EXAMINERS' REPORT

April 2009

Introduction

The attached subject report has been written by the Principal Examiner with the aim of helping candidates. The questions and comments are based around Core Reading as the interpretation of the syllabus to which the examiners are working. They have however given credit for any alternative approach or interpretation which they consider to be reasonable.

R D Muckart
Chairman of the Board of Examiners

June 2009

Comments

Comments on solutions presented to individual questions for the April 2009 paper are given below.

Q1 Answers to this question were satisfactory. Most candidates realised that graduation by reference to a standard table was potentially appropriate, and that graphical graduation might have to be used as a last resort. Credit was given for sensible points other than those mentioned in the specimen solution below.

Q2 In part (ii) some explanation of the correct possible values of a and b was required for full credit. A common error in part (ii) (a) was to write $0 < a < 1$ and $0 < b < 1$, ignoring the possibility that a and b could equal 1.

Q3 This bookwork question was well answered by many candidates. Credit was given for sensible points other than those mentioned in the specimen solution below.

Q4 Answers to parts (i) and (ii) were generally good, with a substantial proportion of candidates scoring full marks. Part (iii) was much less convincingly answered. Although not all the points mentioned in the specimen solutions below were required for full credit, many candidates only included the briefest of comments, and consequently scored few marks.

Q5 Most candidates simply wrote down the formula for Pr[G=t] (which is given in the book of Formulae and Tables) and then explained what each bracketed expression in the formula meant. Few candidates gave more than the briefest explanation of why the test is useful, and what it is designed to achieve, and still fewer gave any indication of how the test was to be performed.

Q6 Answers to this question were very disappointing. Although this was slightly more demanding than some exposed-to-risk questions in the past, many candidates seemed to have little notion of how to approximate the central exposed to risk.

Q7 This question was generally well answered, although part (ii) was less well answered than similar questions on previous papers in which examples relevant to actuarial work were asked for. Marks were deducted in part (ii) for problems which seemed trivial, or where essentially the same examples were given for more than one class of models.

Q8 Few candidates made a serious attempt at this question. Many answers consisted of an attempt at part (i) followed by a description of the state space in part (ii)(a), the general expression for the Kolmogorov equations, and a statement in part (iii) that the distribution of holding times was exponential. Few candidates attempted to write down the matrix in part (ii). Note that credit was given in part (iv) for errors carried forward from incorrect matrices in part (ii).

Q9 Part (i) of this question was well answered by a good proportion of candidates. Fewer managed to calculate the values of B and c in part (ii), partly due to algebraic errors. Credit was given for the calculation of B to candidates who calculated an incorrect value for c but then correctly computed the value of B which corresponded to their value of c. Part (iii) was poorly answered, with many candidates offering no comments at all.

Q10 Answers to this question were very disappointing. Parts (i) and (ii) were bookwork based on the Core Reading, yet many candidates seemed not to understand what was required. Part (iii) was rather better answered. Candidates who derived an incorrect hazard function in part (iii) could score full credit in part (iv) for correct sketches of these incorrect hazards. Indeed, of the relatively small number of candidates who scored highly for the sketches in part (iv), some did indeed produce correct plots of incorrect (and sometimes much more complicated) hazard functions.

Q11 This question was well answered by many candidates. The only general weaknesses were steps missing in part (i) and the lack of explanation of where the approximate variance came from in part (iii)(b). In part (iv), an encouraging number of candidates realised that the Cox model was an obvious alternative model, though few made any further comments on how it might be applied to the problem mentioned in the question.

Q12 This question was also well answered by the majority of candidates. Many scored full marks on parts (i), (ii) and (iii), and made a good attempt at part (iv). The comments asked for in part (v) were, however, much less convincingly made. In part (iv), several candidates combined the two categories "4 claims" and "5 claims" because the expected value was small. Full credit was given for this if the chi-squared statistic was computed correctly, and the number of degrees of freedom was correct for this alternative. However, candidates who performed the test on the reduced number of categories "0 claims", "1 claim" and "2 or more claims" were penalised.

1 Graduation by reference to a standard table might be appropriate, if a suitable standard table could be found.

However the fact that the company insures non-standard lives makes it unlikely that a suitable standard table would exist.

Graphical graduation might be used if no suitable standard table can be found.

However it is a last resort as it is difficult to obtain results which are smooth and which adhere to the data.

Graduation using a parametric formula is unlikely to be appropriate as the amount of data in this investigation is likely to be small and it is unlikely that the company will want to produce a standard table.

2 (i) A Markov chain is a stochastic process with discrete states operating in discrete time in which the probabilities of moving from one state to another are dependent only on the present state of the process.

EITHER

If the transition probabilities are also independent of time.

OR

If the *l*-step transition probabilities are dependent only on the time lag, the chain is said to be time-homogeneous.

(ii) (a) In this case the chain is irreducible if the transition probability out of each state is non-zero (or, equivalently, if it is possible to reach the other state from both states)

So requires $0 < a \leq 1$ and $0 < b \leq 1$

(b) The chain is only periodic if the chain must alternate between the states.

So $a = 1$ and $b = 1$.

3 **Benefits**

Complex systems with stochastic elements can be studied.

Different future policies or possible actions can be compared.

In models of complex systems we can control the experimental conditions and thus reduce the variance of the results without upsetting their mean values.

Can calibrate to observed data and hence model interdependencies between outcomes.

Often models are the only practicable means of answering actuarial questions.

Systems with a long time-frame can be studied and results obtainedrelatively quickly.

Limitations

Time or cost or resources required for model development.

In a stochastic model, many independent runs of the model are needed to obtain results for a given set of inputs.

Models can look impressive and there is a danger this results in false sense of confidence.

Poor or incredible data input or assumptions will lead to flawed output.

Users need to understand the model and the uses to which it can safely be put — the model is not a "black box".

It is not possible to include all future events in a model (e.g. change in legislation).

Interpreting the results can be a challenge.

Any model will be an approximation.

Models are better for comparing the impact of input variations than for optimising outputs.

4 (i) (a) Under UDD the number of deaths between exact ages 30 and 35 years is half the number of deaths between exact ages 30 and 40 years.

So the number of deaths between exact ages 30 and 35 years is

½(98,617 – 97,952) = 332.5

and $_5q_{30} = \dfrac{332.5}{98,617} = 0.0033716$.

(b) Let the constant force of mortality be μ.

Then, since $_t p_x = \exp\left(-\int_0^t \mu_{x+s} ds\right)$,

$_{10} p_{30} = \exp(-10\mu)$

so

$$\mu = \frac{-\log_e\left(_{10} p_{30}\right)}{10} = \frac{-\log_e\left(97,952/98,617\right)}{10} = 0.0006766 .$$

$_5q_{30} = 1 - {}_5 p_{30} = 1 - \exp(-5\mu)$

$= 1 - \exp[(-5)(0.0006766)] = 0.0033773$.

(ii) EITHER

The number of survivors to exact age 35 years is

$98,617\,{}_5 p_{30} = 98,617(1 - {}_5q_{30})$,

so for UDD this is

$98,617(1 - 0.0033716) = 98,284.5$,

and under a constant force of mortality this is

$98,617(1 - 0.0033773) = 98,283.9$.

OR

Under UDD the number of survivors to exact age 35 years is (98,617 + 97,952)/2 = 98,284.5.

Under a constant force of mortality the number of survivors to exact age 35 years is given by

$$\sqrt{98,617 * 97,952} = 98,283.9$$

(iii) The actual number of survivors to exact age 35 years is higher (or, equivalently, mortality is lighter) than that under either the UDD or the constant force assumptions.

The actual number of survivors implies that there were 258 deaths between ages 30 and 35 years and 407 deaths between ages 35 and 40 years.

The actual data reveal that the force of mortality is higher between ages 35 and 40 years than it is between ages 30 and 35 years for females in English Life Table 15, which suggests that the force of mortality is increasing over this age range.

The assumption of UDD implies an increasing force of mortality.

The actual force of mortality seems to be increasing even faster than is implied by UDD.

A constant force of mortality is unlikely to be realistic for this age range.

Used over a 10-year age span the assumption of UDD is unlikely to be appropriate, whereas used over single years of age it is acceptable.

5 Suppose we have a set of n crude mortality rates for a given age range x to $x + n - 1$, and we wish to compare them to a standard set of n mortality rates for the same age range.

If the mortality underlying the crude rates is the same as that of the standard set of rates (the null hypothesis), then we should expect the difference between the two sets of rates to be due only to sampling variability.

The grouping of signs test tests the null hypothesis by examining the number of groups of consecutive positive deviations among the n ages, where a positive deviation occurs when the crude rate exceeds the corresponding rate in the standard set.

Suppose there are a total of m positive deviations, $n - m$ negative deviations and G positive groups.

Then the number of possible ways to arrange t positive groups among $n - m$ negative deviations is $\binom{n-m+1}{t}$.

There are $\binom{m-1}{t-1}$ ways to arrange *m* positive signs into *t* positive groups.

There are $\binom{n}{m}$ ways to arrange *m* positive and *n* – *m* negative signs.

Therefore the probability of exactly *t* positive groups is

$$\Pr[G=t]=\frac{\binom{n-m+1}{t}\binom{m-1}{t-1}}{\binom{n}{m}}$$

The grouping of signs test then evaluates $\Pr[t \le G]$ under the null hypothesis.

If this is less than 0.05 we reject the null hypothesis at the 5% level.

6 (i) (a) The relevant recovery rates can be estimated as

$$r_x = \frac{d_x}{E_x^c}, \quad x = 0, 1, 2, \ldots \text{ months}$$

where d_x is the number of persons recovering in the calendar month that was *x* months after the calendar month of their operation, and E_x^c is the central exposed to risk.

(b) We need to ensure that the E_x^c correspond to the data on persons recovering

The hospital's data imply a calendar month rate interval for the recoveries, running from the first day of each month until the last day of each month.

Using the monthly "census" data, a definition of E_x^c which corresponds to the deaths data can be obtained as follows.

We observe $P_{x,t}$ = number of lives under observation for whom the time elapsing since the operation was between $x - ½$ and $x + ½$ months, where *t* is the time in months since 1 January 2008.

Therefore, using the census formula:

$$E_x^c = \int_0^{12} P^*_{x,t}\, dt = \sum_0^{11} \tfrac{1}{2}\left(P^*_{x,t} + P^*_{x+1,t+1}\right),$$

where $P^*_{x,t} = \tfrac{1}{2}(P_{x-1,t} + P_{x,t})$.

We assume all months are the same length, and that the numbers in the hospital vary linearly across each month.

(ii) At the start of the rate interval, durations since the operation range from $x - 1$ to x months, so the average duration is $x - ½$, assuming operations take place evenly across the month.

r_x estimates the recovery rate at the mid-point of the rate interval.

This is exactly x months since the operation, so $f = 0$.

7 (i) Processes can be classified, first, according to whether their state space (i.e. the range of states they can possibly occupy) is discrete or continuous

For processes operating in both discrete and continuous state space the time domain can either be discrete or continuous

Therefore we have four possible types of process

EITHER

2 types of state space × 2 types of time domain

OR

State space	Time domain
Discrete	Discrete
Discrete	Continuous
Continuous	Discrete
Continuous	Continuous

(ii)

Type of process	*Statistical model*	*Problem of relevance to food retailer*	*Problem of relevance to a general insurer*
SS Discrete/ T Discrete	Markov chain Markov jump chain Counting process Random walk	Whether or not particular product out of stock at the end of each day	No claims bonus
SS Discrete/ T Continuous	Counting process Poisson process Markov jump process Compound Poisson process	Rate of arrival of customers in shop	Number of claims received monitored continuously
SS Continuous/ T Discrete	ARIMA time series model General random walk White noise	Value of goods in stock at the end of each day	Total amount insured on a certain type of policy valued at the end of each month
SS Continuous/ T Continuous	Compound Poisson process Brownian motion Ito process	Volume (or value) of trade in shop over a continuous period of time	Value of claims arriving monitored continuously

8 (i) There are x infected cats and hence $10 - x$ uninfected cats.

Flea transmission requires one of the x infected cats to meet one of the $(10 - x)$ uninfected cats.

(ii) The total number of pairings of cats is $\binom{10}{2} = 45$.

So the probability of a meeting resulting in an increase in the number of cats with fleas is $0.5x(10 - x)/45$.

As this depends only on the number of cats currently infected, and meetings occur according to a Poisson process, the number of infected cats over time follows a Markov jump process.

(a) The state space is the number of cats infected {0,1,2,,…..10}

(b) The generator matrix is

$$A = \frac{\mu}{90}\begin{pmatrix} 0 & 0 & & & & & & & & & \\ & -9 & 9 & & & & & & & & \\ & & -16 & 16 & & & & & & & \\ & & & -21 & 21 & & & & & & \\ & & & & -24 & 24 & & & & & \\ & & & & & -25 & 25 & & & & \\ & & & & & & -24 & 24 & & & \\ & & & & & & & -21 & 21 & & \\ & & & & & & & & -16 & 16 & \\ & & & & & & & & & -9 & 9 \\ & & & & & & & & & & 0 \end{pmatrix}$$

Kolmogorov's equations:

EITHER

forward form $\frac{d}{dt}P(t) = P(t)A$

OR

backward form $\frac{d}{dt}P(t) = AP(t)$

(iii) Holding times are exponentially distributed.

With mean $\frac{90}{\mu x(10-x)}$ OR parameter $\frac{\mu x(10-x)}{90}$.

(iv) Total expected time is the sum of the mean holding times.

$$= \frac{90}{\mu}\sum_{x=1}^{9}\frac{1}{x(10-x)} = \frac{90}{\mu}\left(\frac{1}{9}+\frac{1}{16}+\frac{1}{21}+\frac{1}{24}+\frac{1}{25}+\frac{1}{24}+\frac{1}{21}+\frac{1}{16}+\frac{1}{9}\right)$$

= 50.92/μ

9 (i) Under Gompertz's Law

$\mu_x = Bc^x$.

Since

$${}_t p_x = \exp\left(-\int_0^t \mu_{x+w}\,dw\right),$$

we have $${}_t p_x = \exp\left(-\int_0^t Bc^{x+w}\,dw\right) = \exp\left(-\left|\frac{Bc^x c^w}{\ln c}\right|_0^t\right),$$

which is $$\exp\left(-\frac{\left[Bc^x c^t - Bc^x\right]}{\ln c}\right) = \left[\exp\left(\frac{-B}{\ln c}\right)\right]^{c^x(c^t-1)}.$$

(ii) Define $Q = \left[\exp\left(\dfrac{-B}{\ln c}\right)\right]^{c^{50}}$

ln 0.995 = (c – 1) ln Q

ln 0.989 = (c^2 – 1) ln Q

$$\frac{(c^2-1)}{(c-1)} = \frac{(c-1)(c+1)}{(c-1)} = 2.20665$$

c = 1.20665

Therefore Q = 0.976036128

$$\left[\exp\left(\frac{-B}{\ln 1.20665}\right)\right]^{1.20665^{50}} = 0.976036128$$

$B = 3.797 * 10^{-7}$.

(iii) In this example, only two observations are provided so there is an analytical solution to the Gompertz model.

This is unrealistic as in general a graduation process would be used to provide a fit to a set of crude rates.

This could be done by weighted least squares or maximum likelihood.

The more general graduation process allows the fitting of more complex models from the Gompertz-Makeham family which have the form

μ_x = polynomial(1) + exp(polynomial(2))

the parameters of which cannot always so easily be estimated by the method used in part (ii).

10 (i) (a) $S_x(t) = \Pr[T_x > t]$

(b) EITHER

Since $\Pr[T_x > t] = \Pr[T > x+t \mid T > x] = \dfrac{\Pr[T > x+t]}{\Pr[T > x]}$

and $S(t) = \Pr[T > t]$,

then $S_x(t) = \dfrac{S(x+t)}{S(x)}$.

OR

Since $S_x(t) = {}_t p_x$, then using the consistency principle ${}_{x+t}p_0 = {}_t p_x \cdot {}_x p_0$

Therefore ${}_t p_x = S_x(t) = \dfrac{{}_{x+t}p_0}{{}_x p_0} = \dfrac{S(x+t)}{S(x)}$.

(ii) EITHER

$$\mu_{x+t} = -\frac{1}{\Pr[T_x > t]}\frac{d}{dt}[\Pr(T_x > t)]$$

OR

$$\mu_{x+t} = \lim_{h \to 0^+} \frac{1}{h}\left(\Pr[T_x \le t+h \mid T_x > t\right)$$

(iii) EITHER

If the density function of T_x is $f_x(t)$, then we can write

$$f_x(t) = S_x(t)\mu_{x+t} = -\frac{d}{dt}S_x(t)$$

Therefore $\mu_{x+t} = -\frac{1}{S_x(t)}\frac{d}{dt}S_x(t)$

If $S_x(t) = \exp\left(-(\lambda t)^\beta\right)$, therefore, we have

$$\mu_{x+t} = -\frac{1}{\exp\left(-(\lambda t)^\beta\right)}\frac{d}{dt}\exp\left(-(\lambda t)^\beta\right)$$

$$\mu_{x+t} = -\frac{1}{\exp\left(-(\lambda t)^\beta\right)}\left(\exp\left(-(\lambda t)^\beta\right)\right)\left(-\lambda^\beta \beta t^{\beta-1}\right) = \lambda^\beta \beta t^{\beta-1}$$

OR

$$S_x(t) = \exp\left[-\int_0^t \mu_{x+s}ds\right] = \exp\left[-(\lambda t)^\beta\right].$$

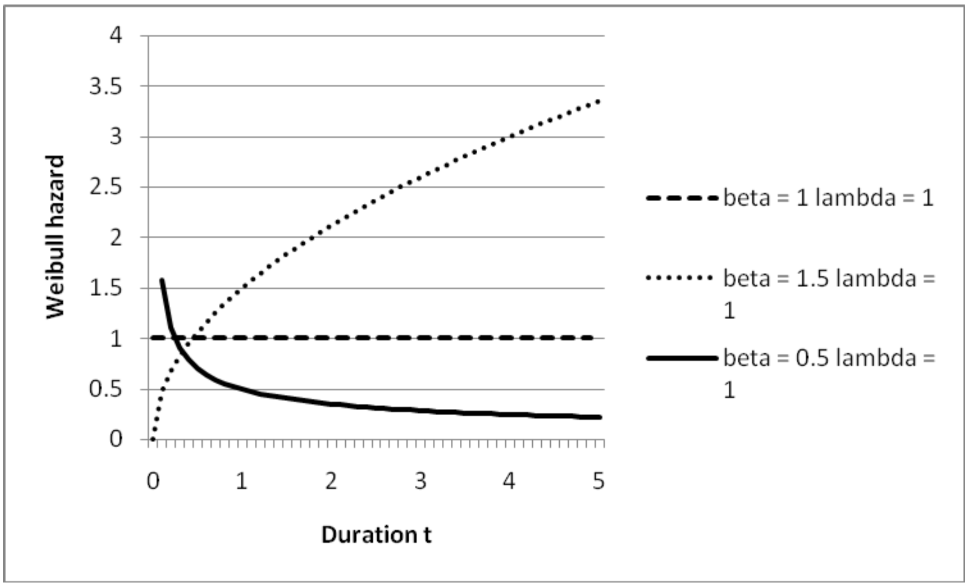
So

$$\frac{d}{dt}\left[\int_0^t \mu_{x+s}ds\right] = \mu_{x+t} = \frac{d}{dt}\left[(\lambda t)^\beta\right],$$

and hence

$$\mu_{x+t} = \beta\lambda^\beta t^{\beta-1}.$$

(iv)



11 (i) Condition on the state occupied at t.

We have

$${}_{t+dt}p_x^{12} = {}_t p_x^{11}\,{}_{dt}p_{x+t}^{12} + {}_t p_x^{12}\,{}_{dt}p_{x+t}^{22}.$$

since it is impossible to leave states 3 and 4 once entered.

Also, ${}_{dt}p_{x+t}^{22} = 1$,

since state 2 is an absorbing state.

We now assume that, for small dt,

$${}_{dt}p_{x+t}^{12} = \mu_{x+t}^{12}dt + o(dt)$$

where $o(dt)$ is the probability that a life makes two or more transitions in the time interval dt, and

$$\lim_{dt\to 0}\frac{o(dt)}{dt} = 0.$$

Substituting for ${}_{dt}p_{x+t}^{12}$ gives

$${}_{t+dt}p_x^{12} = \mu_{x+t}^{12}\,{}_t p_x^{11}dt + {}_t p_x^{12} + o(dt)$$

Thus

$${}_{t+dt}p_x^{12} - {}_t p_x^{12} = \mu_{x+t}^{12}\,{}_t p_x^{11}dt + o(dt)$$

and

$$\frac{\partial}{\partial t}\,{}_t p_x^{12} = \lim_{dt\to 0^+}\frac{{}_{t+dt}p_x^{12} - {}_t p_x^{12}}{dt} = \mu_{x+t}^{12}\,{}_t p_x^{11}$$

(ii) (a) Suppose we observe d^{12} deaths from heart disease, d^{13} deaths from cancer and d^{14} deaths from other causes.

Suppose also that we observe the waiting time for each life, and that the total observed waiting time is V, being the sum of the waiting times for each life.

Then the likelihood of the data is given by

$$L \propto \exp\left[-\left(\mu^{12}+\mu^{13}+\mu^{14}\right)V\right](\mu^{12})^{d^{12}}(\mu^{13})^{d^{13}}(\mu^{14})^{d^{14}}.$$

(b) The maximum likelihood estimator of μ^{12} isobtained by differentiating this expression (or its logarithm) with respect to μ^{12} and setting the derivative equal to zero.

Taking logarithms produces

$$\log L = -(\mu^{12}+\mu^{13}+\mu^{14})V + d^{12}\log\mu^{12} + d^{13}\log\mu^{13} + d^{14}\log\mu^{14} + K$$

(where K is a constant)

Partially differentiating this with respect to μ^{12} leads to

$$\frac{\partial \log L}{\partial \mu^{12}} = -V + \frac{d^{12}}{\mu^{12}},$$

and setting the partial derivative equal to zero leads to the solution

$$\hat{\mu}^{12} = \frac{d^{12}}{V}.$$

Since $\frac{\partial^2 \log L}{(\partial\mu^{12})^2} = -\frac{d^{12}}{(\mu^{12})^2}$, the second derivative is always negative and so we have a maximum.

(iii) (a) The maximum likelihood estimate of the force of mortality from heart disease is 34/1,065 = 0.0319249

(b) The variance of the maximum likelihood estimator of μ^{12} is asymptotically $\frac{\mu^{12}}{E[V]}$, where $E[V]$ is the expected waiting time in the state "alive" and μ^{12} is the "true" population value of the force of mortality from heart disease.

This may be approximated by using the observed force of mortality and the observed waiting time, so that an estimate of the variance is

$$\frac{0.0319249}{1{,}065} = 0.000029976.$$

The estimated standard error is therefore

$\sqrt{0.000029976} = 0.00547507$.

The 95% confidence interval is therefore

$$0.0319249 \pm (1.96)0.00547507 = 0.0319249 \pm 0.0107311$$
$$= (0.0212, 0.0427).$$

(iv) Using the four state model, the lives in the investigation would have to be stratified according to the risk factors and the transition intensities estimated separately for each stratum.

This is likely to run into problems of small numbers.

Using a Cox regression model with death from heart disease as the event of interest and the risk factors as covariates would avoid this problem.

Lives who died from other causes could be treated as censored at the durations when they died.

12 (i) The probability of making the relevant number of claims is:

P[0 claims] = exp(–0.3) = 0.740818
P[1 claim] = 0.3exp(–0.3) = 0.222245

So *P*[2 or more claims] = 1 − 0.740818 − 0.222245 = 0.036936

Therefore the transition matrix *P* is given by:

$$\begin{pmatrix} 0.259182 & 0.740818 & 0 & 0 \\ 0.259182 & 0 & 0.740818 & 0 \\ 0.036936 & 0.222245 & 0 & 0.740818 \\ 0 & 0.036936 & 0.222245 & 0.740818 \end{pmatrix}$$

(ii) $\pi = \pi P$

$$\pi_1 = 0.259182\pi_1 + 0.259182\pi_2 + 0.036936\pi_3 \qquad (1)$$
$$\pi_2 = 0.740818\pi_1 + 0.222245\pi_3 + 0.036936\pi_4 \qquad (2)$$
$$\pi_3 = 0.740818\pi_2 + 0.222245\pi_4 \qquad (3)$$
$$\pi_4 = 0.740818\pi_3 + 0.740818\pi_4 \qquad (4)$$
$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$$

Using (4)

$\pi_3 = [(1 - 0.740818) / 0.740818] * \pi_4 = 0.349859\pi_4$.

In (3)

$\pi_2 = [(0.349859 - 0.222245) / 0.740818] * \pi_4 = 0.17226\pi_4$.

Then in (2)

$\pi_1 = [(0.17226 - 0.036936 - 0.222245 * 0.349859) / 0.740818] * \pi_4 = 0.07771\pi_4$

So

$\pi_4 = 1 / (1+0.349859+0.17226+0.07771)=0.625067$
$\pi_3 = 0.218685$
$\pi_2 = 0.107674$
$\pi_1 = 0.048574$

(iii) Average discount =

60%*0.625067+50%*0.218685+25%*0.107674 = 51.13%

(iv) The total number of policyholders shown is 130,200.

Number of claims	*Probability*	*Expected Number*	*Observed*	*$(O - E)^2/E$*
0	0.740818221	96454.53	96632	0.327
1	0.222245466	28936.35	28648	2.873
2	0.03333682	4340.45	4400	0.817
3	0.003333682	434.05	476	4.054
4	0.000250026	32.55	36	0.366
5	1.50016*E*–05	1.95	8	18.771

Null hypothesis: the data come from a source where the underlying distribution of number of claims follows a Poisson distribution with mean 0.30.

The test statistic $z = \sum_i (O_i - E_i)^2 \big/ E_i$ is distributed as chi-square with (6 – 1(parameter) – 5 degrees of freedom under the null hypothesis.

This is a one-tailed test, and the upper 5% point of the chi-squared distribution with 5 degrees of freedom is 11.07.

The observed value of the test statistic is 27.2.

As 27.2 > 11.07 we reject the null hypothesis.

(v) As the goodness of test fails, the discount level calculated assuming the Poisson distribution may be incorrect.

The goodness-of-fit test fails due to a larger number of multiple claims than expected.

Conversely a higher number of policyholders make no claims than expected (within the mean of 0.30), so the average discount level may be understated.

The average discount level calculated from the data could usefully be compared with that estimated using the Poisson distribution.

END OF EXAMINERS' REPORT